

<b>15th International Roundtable on Business Survey Frames</b>
<b>Washington, D.C. – October 22 – 26, 2001</b>
<p style="text-align: right;"><b>Session No 6</b>  <b>Paper No 2</b>  <b>Philippe Brion, Insee, France</b></p>
<b>Use of Administrative Data for Structural Business Surveys in France</b>

The French statistical system has been using administrative data for structural business surveys for a long time, and for different reasons. First for getting some "individual" data (for example by using fiscal data), second for problems relative to the sampling frame, and for the production of estimations (correct weighting of the "individual" data of the survey).

This paper presents the way administrative data are used for structural business surveys in France (and especially for the annual enterprise survey), and also gives examples of problems raised by this practice.

## **1. The annual enterprise surveys in France**

11. These surveys are conducted every year for six broad economic sectors : goods-producing industries, food industries, trade (wholesaling/retailing), services, construction, and transportation, and managed by six different departments. For a global size of 2 millions enterprises, approximately 250000 are surveyed by mail. For the big enterprises, the survey is exhaustive, and under a given size, the selection of the sample is made using a stratification (mixing the region, the number of employees, and the sector determined by the principal economic activity code).

For industry and food industries, the survey is made only for enterprises having more than 20 employees (and exhaustive). For the non-exhaustive part of the other sectors, one-half of the sampled group is replaced every year.

12. Administrative data are used for the annual enterprise surveys for two reasons :

- the question of weighting the sample units, particularly because of non responses : are these non responses "dead" units, or not ? Having some external information in administrative sources may be useful for this purpose (see part 4 of this paper) ;
- for some big units non-responses, it is dangerous to use standard methods of extrapolation, (using the data of the responses) ; it is better to get fiscal data (**annual income statements**) and to use them, because of the specificity of these units.

Since 1996, the use of annual income statements has also been introduced for small enterprises (see part 2).

13. One may wonder if it would be possible to use administrative data, especially fiscal data, alone. But the variables available in the fiscal source are not as complete as those available in the survey : especially, the principal economic activity code (called APE code) is known, in the fiscal source, as of poor quality. That is the reason why this latter source is only used as an additional information.

14. The use of administrative data is made easy, in France, by the existence of the business register, designed for both statistical and administrative purposes : so, legal, tax and social-insurance agencies use the same register identification number SIREN for enterprises.

## **2. For small enterprises, a specific device**

For the small units of wholesaling/retailing, services, construction and transportation, statisticians use a specific device in order to reduce the burden enterprises have to support. For these categories, the survey is not conducted every year, but only every two years, and the other year the fiscal data are used, for the sample which was surveyed the previous year. That is called "small enterprises strategy".

In order to smooth the quantity of work of the departments in charge of services, trade

(wholesaling/retailing), transportation and construction, an alternation has been introduced : for example, when the survey is made for small enterprises of wholesaling, fiscal data are used for retailing ; next year, the survey is made for retailing, and fiscal data are used for wholesaling. For the annual enterprise surveys conducted in 2001 (concerning the data of year 2000), the “small enterprise strategy” concerned wholesaling and business services (table 1).

Table 1: “small enterprises strategy” for the 2000 and 2001 annual enterprise surveys

2001

Economic sector	Number of units using questionnaires	Number of units using fiscal data
Wholesaling	8473	15932
Retailing	39927	0
Business services	11736	33953
Household services	38959	0

2000

Economic sector	Number of units using questionnaires	Number of units using fiscal data
Wholesaling	25990	0
Retailing	8322	32418
Business services	45831	0
Household services	8788	33326

### 3. Problems raised by the use of individual data

#### 31. Problems raised by the identification number

The number SIREN of the register is used by the fiscal administration, but there are some problems of units missing, and also two or more units having the same identification code. The last case is often due to units having two fiscal declarations (or more) for the same year, for example in case of restructuring.

#### 32. The quality of the fiscal data

Another problem is due to the fact that some variables are not considered as important by the fiscal administration : for example, the fiscal activity code (APE code) is of poor quality. That is the reason why, when using the fiscal data instead of conducting a survey for small enterprises (small enterprises strategy), it is better to use the value of APE code obtained in the survey of the previous year : so, it is necessary to maintain a survey ; the value of the APE code does not change between the year of the survey and the “fiscal” year.

Some data are also less complete, in fiscal files : for example those ones concerning the employees, for which the fiscal source doesn't give the details for part-time employees, the numbers for each quarter, the non salaried, ... The values of these variables have to be estimated using distributions coming from other sources.

#### 33. Problems raised by the production process existing for the business surveys

The annual enterprise surveys managed by the six different departments use common statistical methods, and a software data-editing program has been developed for them. Particularly, some controls, based on ratios, give warnings when the ratio is outside a given interval. When the data are produced by a survey, people treating them may have a contact with the enterprise, by telephone, to check the value of the variables. But when the data are obtained using the fiscal source, the statistical departments don't call enterprises : so, for some kinds of controls, they are sometimes led to widen the intervals, giving less efficiency to them.

Also, the parameters used in the controls are initialized with information coming from the survey of the

previous year. Then, when we use the fiscal data, the first enterprises being controlled are the “births” of small enterprises, for which the survey is still made sending a questionnaire ; the parameters of the controls are not quite well fitted for that category of enterprises, and they will be fitted again after taking into account these first data ; and when the data coming from the fiscal source will be available (with most of small enterprises), the same problem will occur.

#### **4. Using administrative data for the production of estimations (aggregates)**

##### **41. The treatment of non responses**

The problem is to know if these non responses to the annual enterprise surveys are dead enterprises, or not. For that purpose, the **VAT (value added tax) file** is used : it is possible to see, among enterprises, which ones are active, among the non-responses. For the other ones (those for which the VAT file does not give information), an investigation (“administrative” survey) is conducted by people in charge of the register. However, these investigations may produce relatively complete results very late, too late for the results of the annual enterprise surveys (these investigations are useful both for these surveys, and, of course, for the register). So, partial results of these investigations are used, in December, for producing the results of the annual enterprise surveys ; and for the non responses for which no information is available, a “hotdeck” method is applied, using the population of enterprises for which the information “active” or “non-active” is available.

When using the “small enterprises strategy”, one gets few information about non-active enterprises with the fiscal data ; the number of dead enterprises (which is obtained using the “hotdeck method”) is probably underestimated.

##### **42. The restructuring of enterprises**

Different sources are used to find the restructuring of enterprises ; the surveys of the previous year, but also **legal notices**, especially the BALO (Bulletin d’Annonces Légales Obligatoires), an official publication in which listed companies and their subsidiaries are required to advertise their intended full mergers, splits, or partial mergers. The information on restructuring of enterprises let make the calculations of economic evolutions possible (evolutions calculated with constant “outlines” of the enterprises).

##### **43. A specific problem of organization : the “transfers” from one survey department to another one**

When conducting the annual enterprise survey, some units are observed as belonging to another sector than their initial sector : for example, the “industry” department sent a questionnaire, and the unit is now belonging to the wholesaling/retailing sector. For the surveys conducted in 2000 relative to the year 1999, such 1700 transfers were observed ; but only 280 were completely taken into account by the “receiving” department, due to the fact that this department has to send a new questionnaire to the unit, to complete the information collected by the “origin” department (the questionnaires have a common and specific parts). Using fiscal data may be a solution for the “receiving” department, when it is too late to get the data by questionnaire.

##### **44. The problem of “micro-enterprises”**

For these units (approximately 280000 units), the fiscal administration does not produce data files, because their annual income statements are very simplified. When using the “small enterprises strategy”, one has to be very careful with these units : extrapolating them using other units for which fiscal data are available could be dangerous, because the latter are bigger. The treatment of these “micro-enterprises” uses the data of the previous year (the year of the survey), and a growth rate.

It seems quite paradoxical that these enterprises should be the most concerned by the fact of reducing the burden, and that is for them that it is the most difficult to get administrative data.

#### **Conclusion**

Administrative data are very useful for the treatment of the structural business surveys in France.

Getting information from the annual income statements especially (the delivery time of these files by the fiscal administration has been reduced), that reduces the burden on enterprises : for small enterprises, this procedure is used every two years in some economic sectors. But it has some consequences on the organization of the whole treatment of the surveys. And difficulties still exist : for some sectors, the results for two consecutive years, using the “small enterprises strategy” (between the year where the survey is conducted, and the year where the fiscal data are used), present some strange evolutions.

Work has still to be done to better control the way of using this information. Particularly, the fact that we use only one part of the data available (for example, one does not use fiscal data the year where the survey is conducted) might give rise to further developments.

But other administrative data are also used : value added tax files, and legal notices about restructuring ; annual statements of payroll data is a possible source of information about the numbers of employees.

Using information about the distribution of enterprises could also be another possible way of improving the results of these surveys, using methods of re-weighting to adjust the sample to these distributions.